Student: Desiatkin Dmitrii

Group: MS_RO1_Y2.

Instructor: Leon Derczynski

Institution: Innopolis University

Course: Automatic fact verification and fake news detection.

Assignment 3: Automatic detection of misinformation.

Measurements

Problem:

For analysis purpose I have chosen something between measurements and defend techniques from given task list. I especially want to understand the mechanism of fake news creation in social network as well as their spreading. Also it is very important to clarify the set of tools what can be used in potential misinformation campaign.

Analysis:

I decided to start from data that was provided in assignment description, so I have read about Facebook native advertisement dataset.

Also I have downloaded some part of the dataset, and wrote simple parser that helps me extract data from raw pdf format (I have not code it hard, it correctly parses only 60% of information, problem is pdf libraries in Python). But unfortunately I have not found any interesting features in it by simple observing. In my opinion published part is not representative. However it of course contains circumstantial evidence: currency that was used for payment and time of publications.

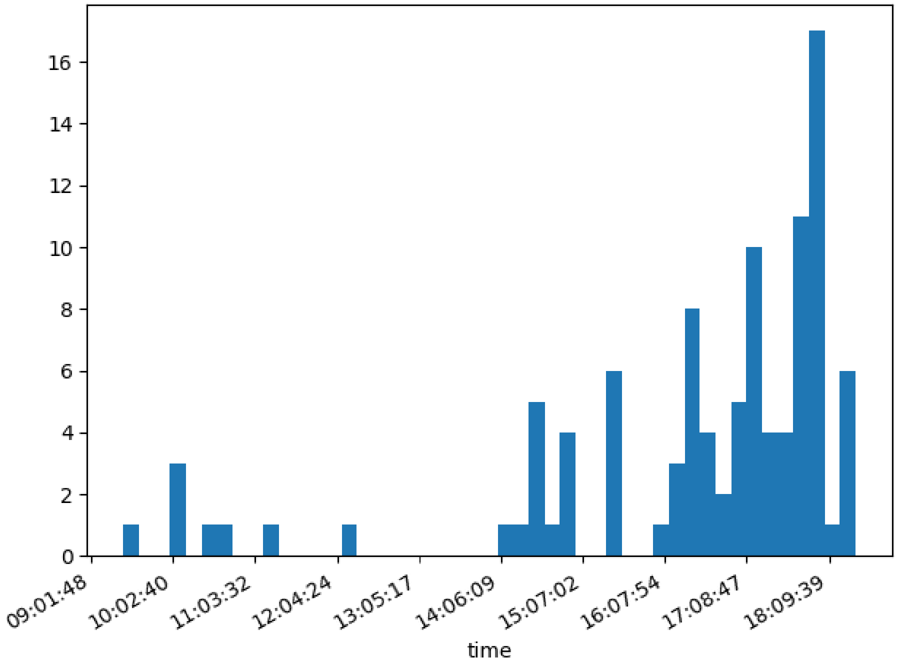
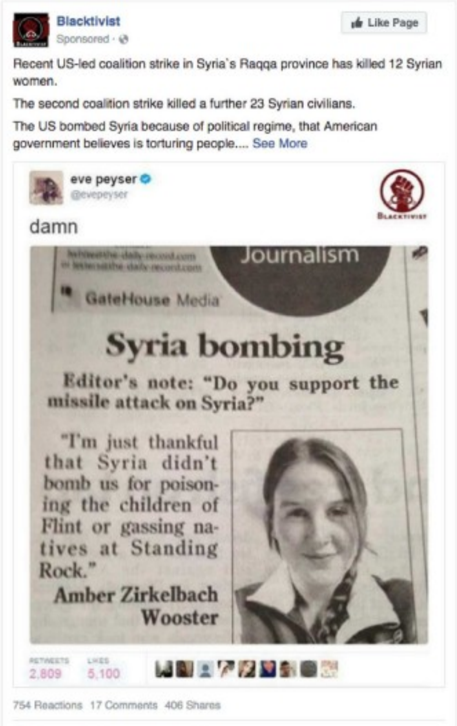


Illustration 1: Time of advertisement creation in Facebook dataset. Here time converted to GMT +3 or simply Moscow time. data-2017-05

Currency that was used for purchasing always was Russian ruble.
Advertisements always have very broad target audience 18 – 65+.
They often contains hot topics, also they have good design (picture/video + meaningful text, can be created in meme style, or use click bait) what increases their virality.



Blacktivist

Sponsored

Recent US-led coalition strike in Syria's Raqqa province has killed 12 Syrian women.

The second coalition strike killed a further 23 Syrian civilians.

The US bombed Syria because of political regime, that American government believes is torturing people.... See More

eve peyser @evepeyser

damn

Syria bombing

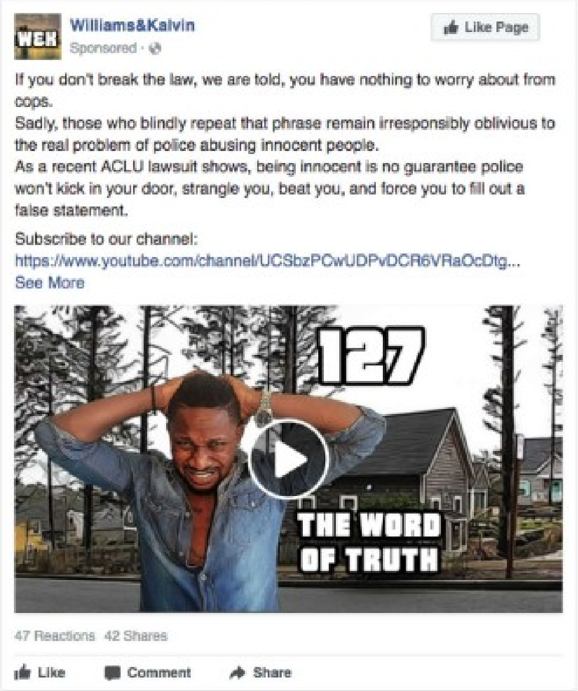
Editor's note: "Do you support the missile attack on Syria?"

"I'm just thankful that Syria didn't bomb us for poisoning the children of Flint or gassing natives at Standing Rock."

Amber Zirkelbach
Wooster

RETWEETS 2,809 LIKES 5,100

754 Reactions 17 Comments 406 Shares



Williams&Kalvin

Sponsored

If you don't break the law, we are told, you have nothing to worry about from cops.
Sadly, those who blindly repeat that phrase remain irresponsibly oblivious to the real problem of police abusing innocent people.
As a recent ACLU lawsuit shows, being innocent is no guarantee police won't kick in your door, strangle you, beat you, and force you to fill out a false statement.

Subscribe to our channel:
https://www.youtube.com/channel/UCSbzPCwUDPvDCR6VRaOcDtg...
See More

47 Reactions 42 Shares

Table 1: Example of context advertisement that was marked as "produced by IRA or affiliated individuals and organizations"

Official H. P. S. C. I. statement about Facebook data:

- 3,393 advertisements purchased (a total 3,519 advertisements total were released after more were identified by the company);
- More than 11.4 million American users exposed to those advertisements;
- 470 IRA-created Facebook pages;
- 80,000 pieces of organic content created by those pages;
- and exposure of organic content to more than 126 million Americans.

However I could not find the methodology that authors used for creation of that dataset. All that was said is that advertisements was manually selected. I have also faced with the same issue about tweeter dataset. The only document that I still need to research is PERMANENT SELECT COMMITTEE ON INTELLIGENCE Minority views (unclassified part), but it contains 100 pages what is pretty large corpus of information for that assignment.
Information of Tweeter dataset also lacks clear methodology of gathering method.

My assumptions is that researchers firstly have studied all advertisements that was paid in Russian currency. I suppose that piece of a data comes directly from Facebook itself. Then they manually mine the set, and picked some number of advertisements that they consider as troll ones. After they have done it they have clustered that set, based on “Ad Creation Time” and “Ad Targeting fields”. After they have done it.

They may further increase the quality of set by manual filtering. I highly doubt that they use some image analysis tools. Because content of pictures are highly uncorrelated within published dataset.

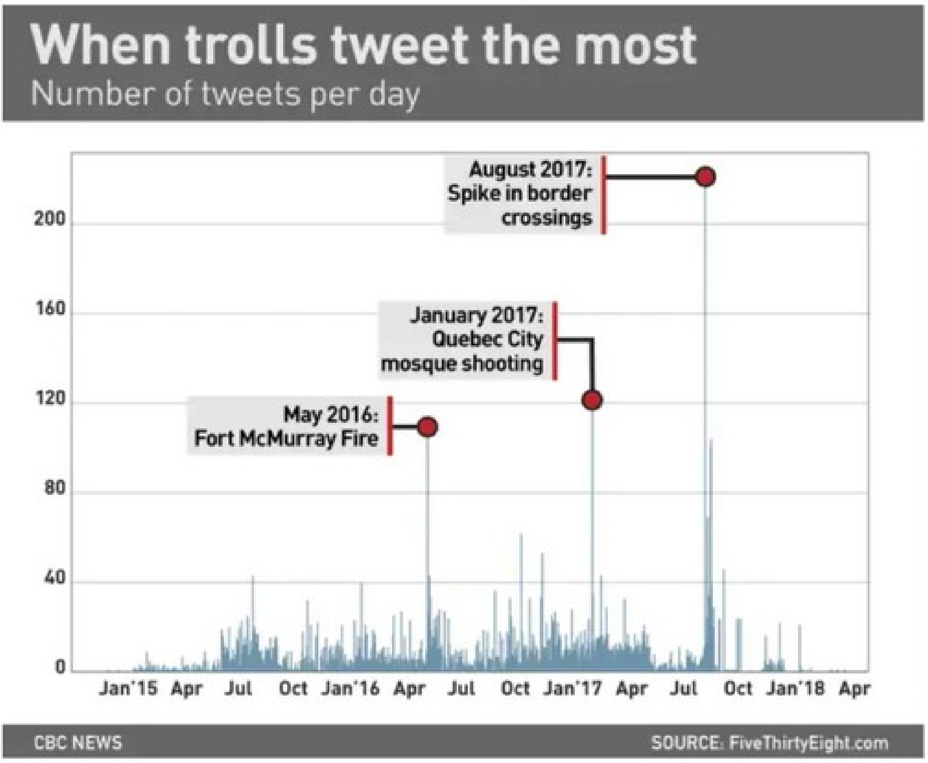
In general I can admit that some system that can distinguish troll posts automatically can be created. Modern neural networks such as GPT2 or BERT showed us that there are exists some complex structures that can analyze text and get some really high level abstractions in it. However we still lacks high level abstractions in computer vision, and it is the crucial part for robust outputs of such systems. Cause in good meme picture contains more than half of the joke, moreover it is become joke only after we read the text. What means that picture can code a lot of meanings, and text help us to choose particular meaning.

Now lets look on Tweeter dataset. As you can see that data is pretty hard to analyze by itself. However other people have already studied some parts of that set (paper, fivethirtyeight).
After reading the paper methodology of data gathering become clear. However they still have used some industrial programs what hardly can be used in academia. Salesforce’s Social Studio to parse individual tweets from list published by Intelligence Committee. But even with that software they have parsed excess tweets, so they need to filter them. After they have started deep qualitative analysis. Firstly they have filtered tweets that in their opinion connected to IRA activity. After they split them on five classes, namely: “Non – English”, “Left Trolls”, “Right Trolls”, “Fearmongers”, “HashtagGamers”.

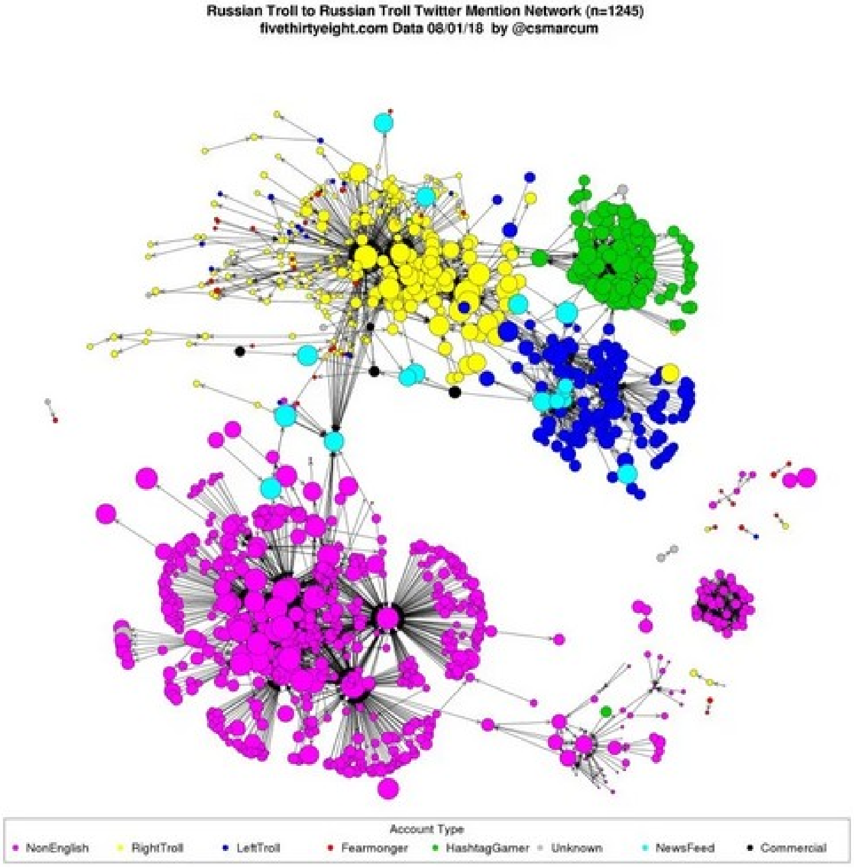
Researchers give good statistical metrics for that division, so is seems to be close to reality. However it is still can be lucky correspondence. Cause as they stated, they do not know particular tactics or goals of IRA. As far as I understand they also have shifted analysis towards text data. So it seems that today does not exist any reliable technique for finding misinformation with image data.

Class “Non – English” have not been analyzed by researchers. In other classes they use next types of features:

1. linguistic,
2. psychological,
3. account name,
4. tweet frequency,
5. tweet peaks (and their connections to events).



After researchers from fivethirtyeight concluded deep study of network relations of troll accounts. They mined the data in order to find retweets between IRA affiliated imposers. It resulted in next graph.


Russian Troll to Russian Troll Twitter Mention Network (n=1245)
fivethirtyeight.com Data 08/01/18 by @csmarcum
Account Type
NonEnglish
RightTroll
LeftTroll
Fearmonger
HashtagGamer
Unknown
NewsFeed
Commercial

It is clear that network is pretty tight. In English segment we have only one small isolated segment. This is also can be considered as circumstantial evidence of some sort of operation. However, it is also possible that here we see confirmation bias.

Conclusion:

During that assignment I have seen how methods that we have studied in the course apply to real data analysis. Also I have not studied that question before so it was very interesting. I still not sure about whether such actions could affect elections in the country with 327,2 millions of people, but observing all that circumstantial evidences, gave me the feeling that IRA could actually exist and perform such operations. Moreover it will be interesting to analyze Russian segment and understand the course of state propaganda. Unfortunately I have not enough time to do it by myself, because of exams, but I hope that I will do it when my study end.

Sources:

https://intelligence.house.gov/social-media-content/

https://intelligence.house.gov/uploadedfiles/exhibit_b.pdf

https://www.aclweb.org/anthology/C18-1287.pdf

https://www.scientificamerican.com/article/biases-make-people-vulnerable-to-misinformation-spread-by-social-media/

https://docs.house.gov/meetings/IG/IG00/20180322/108023/HRPT-115-2.pdf